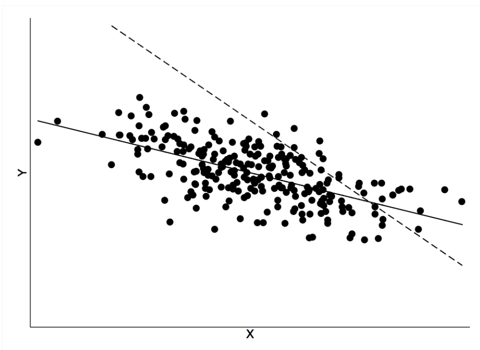
ØAMET2200 · Fall 2019
Worksheet 6

Instructor: Fenella Carpena

October 4, 2019

Exercise 1 Consider the population regression line $y = \beta_0 + \beta_1 x + \epsilon$ and the sample regression line $\widehat{y} = b_0 + b_1 x + e$. Note that for the sample regression line, $e = y - \widehat{y}$. Explain the difference between b_1 and β_1; between the residual e and the population regression error ϵ.

Exercise 2 Consider the population regression line $y = \beta_0 + \beta_1 x + \epsilon$. The figure below shows data from a sample of 250 observations of x and y. One of the lines is the sample regression line, $b_0 + b_1 x$; the other is the population regression line $\beta_0 + \beta_1 x$. Is the sample regression line solid or dashed? Explain.



Exercise 3 Suppose a fire insurance company wants to relate the amount of fire damage in major residential fires to the distance between the burning house and the nearest fire station. The company collected data from a sample of 15 recent fires in a large suburb of a major city. For each fire, the data records the amount of damage (in thousands of kroner), as well as the distance between the fire and the nearest fire station (in kilometers). The lowest distance recorded in the data is 1.8 kilometers, while the highest distance is 6.1 kilometers.

The table below shows the results for the regression of the amount of damage on distance. Assume that all SRM assumptions are satisfied.

Regression Statistics	
r^2	0.923
s_e	23.16
n	15.000

	Coefficients	Standard Error
Intercept	102.78	14.20
Distance	49.19	3.93

(a) Interpret the intercept and slope of the above regression.

(b) Construct a 95% confidence interval for the slope β_1. Interpret this interval.

(c) Use a t-statistic to test the null hypothesis that every 1 kilometer increase in distance between the fire and the fire station is associated with an increase in damages of kr 60,000.

(d) Calculate the p-value for the same hypothesis test given in part (c). Do you reject the null at the 5% level?

(e) Suppose instead that we have the following null and alternative hypotheses: $H_0 : \beta_1 = 0$, and $H_a : \beta_1 \neq 0$. Carry out this test at the 1% significance level using a confidence interval, t-stat, and p-value. Verify that all three methods yield the same conclusion.

(f) Suppose that the insurance company wants to predict the fire damage if a major residential fire were to occur 3.5 kilometers from the nearest fire station. Construct a 95% prediction interval, and interpret it in words.

(g) Suppose instead that the company wants to predict fire damage for homes 8 kilometers away from the fire station. Will the prediction interval be reliable? Explain why or why not.

Exercise 4 Mark each statement as true or false and explain why.

(a) To estimate the effect of advertising on sale using a regression, you should look for periods with steady levels of advertising rather than periods in which advertising varies.

(b) Doubling the sample size used to fit a regression can be expected to reduce the standard error of the slope by about 30%

(c) The assumption of a normal distribution of the errors in a regression model is important for the confidence interval of the slope.

(d) If the SRM is used to model data that do not have constant variance, then 95% prediction intervals produced by this model are longer than needed.

(e) The presence of an outlier in the data used to fit a regression causes the estimated model to have a lower r^2 than it should.

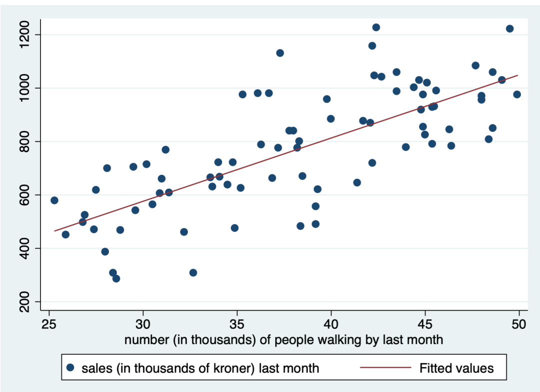
Exercise 5 (Baker Hansen) You are in charge of deciding the location of a new Baker Hansen store. You are considering a variety of different possible locations with different foot traffic levels. You have (fictional) data for 80 Baker Hansen locations. For each location, you have the following two variables from last month: (1) `sales`, total sales in thousands of kroner; (2) `foot_traffic`: number of people walking by the store in thousands. Assume that all locations charge the same price and are located in similar neighborhoods.
You regress `sales` on `foot_traffic` and you obtain the regression results and figures below.

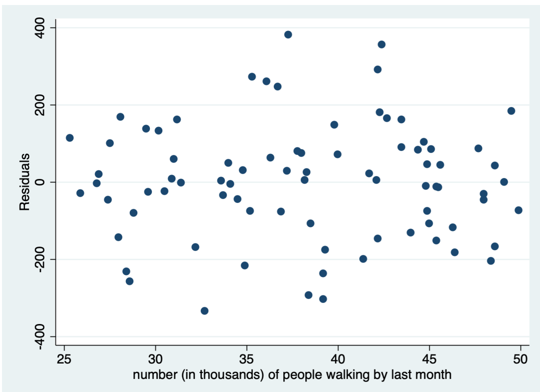
```
. regress sales foot_traffic ;
```

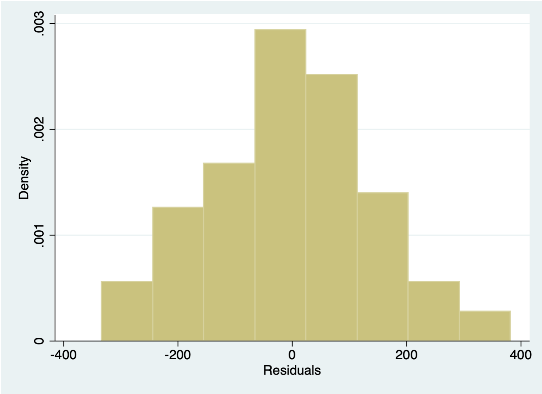
Source	SS	df	MS
Model	2148270.28	1	2148270.28
Residual	1767674.71	78	22662.4963
Total	3915944.99	79	49568.9239

Number of obs =	80
F(1, 78) =	94.79
Prob > F =	0.0000
R-squared =	0.5486
Adj R-squared =	0.5428
Root MSE =	150.54

sales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
foot_traffic	23.67286	2.431421	9.74	0.000	18.83228	28.51345
_cons	-133.8097	94.58436	-1.41	0.161	-322.1127	54.49326







(a) Interpret b_0 and b_1.

(b) Do the five SRM conditions hold?

(c) Consider the following situations. Explain whether they violate the SRM of sales and foot traffic. If so, which SRM condition is violated?

(i) Baker Hansen shops located in busier areas (more foot traffic) also have better customer service.

(ii) The variation in sales is lower in Baker Hansen shops located in busier areas.

(iii) The effect of a 1,000 increase in foot traffic on sales is larger in busier areas.

(d) Use a t-statistic to determine if the slope is statistically significantly different from zero at the 5% level.

(e) What is the p-value of the hypothesis test in part (d)?

(f) What is the 95% CI for β_1?

(g) What is the 80% CI for β_0?

(h) What is the 95% prediction interval for the monthly sales of a location with foot traffic level of 50,000?

Exercise 6 (Stata) The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) describes the relationship between returns on a speculative asset (typically returns on stock) and returns on the whole stock market. The underlying theory describes the risk of owning an asset, where risk refers to variation in returns over time. Part of the risk of owning an asset is associated with overall movements in the market as a whole and so is called *market risk*. The remaining *idiosyncratic risk* is unique to the asset. The underlying theory promises that markets will pay investors for taking market risk, but not for assuming idiosyncratic risk, which it treats as a gamble.

According to the CAPM, a simple regression model describes the returns R_t on a stock over some time period as

$$R_t = \alpha + \beta M_t + \epsilon_t$$

where M_t is the return on the market and ϵ_t denotes the effects of other factors on the stock return. The intercept is called the alpha of the stock and the slope is called the beta of the stock. If $\beta = 0$, then movements in the market don't influence the value of the stock. If $\beta = 1$, then the stock tends to move up and down with changes in the market. Stocks withe $\beta > 1$, sometimes called growth stocks, amplify movements in the market whereas those with $\beta < 1$ attenuate swings in the market. The claim of CAPM that investors are not compensated for taking idiosyncratic risks implies that $\alpha = 0$.

Note that in this exercise, parts (a)-(d) are conceptual questions, and parts (e)-(h) will require Stata. The data for this exercise is `capm.xlsx`.

Motivation

(a) Hedge funds often claim that they are able to pick investments for which $\alpha > 0$. What are they claiming about these investments?

(b) Investors often seek stocks that diversify their risk into uncorrelated investments. Should such investors seek stocks with large β or $\beta \approx 0$?

Method

(c) A stock advisor claims that Berkshire Hathaway, the investment company run by Warren Buffett, generates "positive alpha." How can we test this claim using a regression model?

(d) How can we use a confidence interval to test the claim that beta is one for Berkshire Hathaway?

Mechanics

(e) The data for this exercise gives monthly returns on Berkshire Hathaway and the overall stock market since January 1980 (`capm.xlsx`). Use these data to fit the CAPM regression and check whether the SRM is appropriate.

(f) The precise returns used in the CAPM are so-called excess returns, formed by subtracting the return R_f on a risk free asset (usually Treasury Bills) from the return on a risky asset. Form the excess returns on Berkshire Hathaway and the stock market and fit the regression

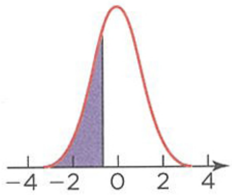
$$(R_t - R_f) = \alpha + \beta(M_t - R_f) + \epsilon_t.$$

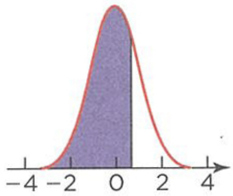
Compare the estimates of the slope and intercept from this equation to those obtained in (e) using the nominal returns. Why are the differences so small?

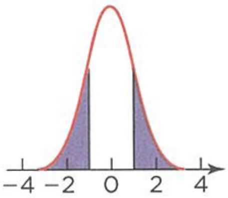
Message

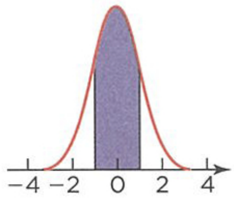
(g) Does Berkshire Hathaway produce excess alpha? Explain.

(h) Is Berkshire Hathaway a growth stock? Explain.



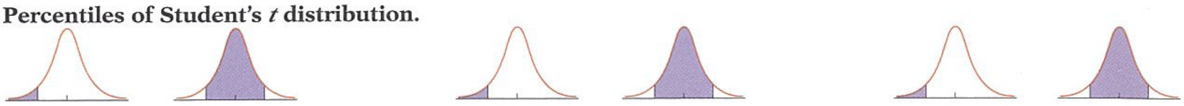






z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\|Z\| > z)$	$P(\|Z\| \leq z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	0.32	0.68	0.64	0.36
0.4959	0.31	0.69	0.62	0.38
0.5244	0.30	0.70	0.60	0.40
0.5828	0.28	0.72	0.56	0.44
0.6433	0.26	0.74	0.52	0.48
0.6745	0.25	0.75	0.50	0.50
0.7063	0.24	0.76	0.48	0.52
0.7388	0.23	0.77	0.46	0.54
0.7722	0.22	0.78	0.44	0.56
0.8064	0.21	0.79	0.42	0.58
0.8416	0.20	0.80	0.40	0.60
0.8779	0.19	0.81	0.38	0.62
0.9154	0.18	0.82	0.36	0.64
0.9542	0.17	0.83	0.34	0.66
0.9945	0.16	0.84	0.32	0.68
1.0364	0.15	0.85	0.30	0.70
1.0803	0.14	0.86	0.28	0.72
1.1264	0.13	0.87	0.26	0.74
1.1750	0.12	0.88	0.24	0.76
1.2265	0.11	0.89	0.22	0.78
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999

T–TABLE **Percentiles of Student's *t* distribution.**



$df = 1$

t	$P(T_1 \le -t)$	$P(-t \le T_1 \le t)$
3.078	0.1	0.8
6.314	0.05	0.9
12.71	0.025	0.95
31.82	0.01	0.98
63.66	0.005	0.99
318.3	0.001	0.998
636.6	0.0005	0.999
6366	0.00005	0.9999

$df = 2$

t	$P(T_2 \le -t)$	$P(-t \le T_2 \le t)$
1.886	0.1	0.8
2.920	0.05	0.9
4.303	0.025	0.95
6.965	0.01	0.98
9.925	0.005	0.99
22.33	0.001	0.998
31.60	0.0005	0.999
99.99	0.00005	0.9999

$df = 3$

t	$P(T_3 \le -t)$	$P(-t \le T_3 \le t)$
1.638	0.1	0.8
2.353	0.05	0.9
3.182	0.025	0.95
4.541	0.01	0.98
5.841	0.005	0.99
10.21	0.001	0.998
12.92	0.0005	0.999
28.00	0.00005	0.9999

$df = 4$

t	$P(T_4 \le -t)$	$P(-t \le T_4 \le t)$
1.533	0.1	0.8
2.132	0.05	0.9
2.776	0.025	0.95
3.747	0.01	0.98
4.604	0.005	0.99
7.173	0.001	0.998
8.610	0.0005	0.999
15.54	0.00005	0.9999

$df = 5$

t	$P(T_5 \le -t)$	$P(-t \le T_5 \le t)$
1.476	0.1	0.8
2.015	0.05	0.9
2.571	0.025	0.95
3.365	0.01	0.98
4.032	0.005	0.99
5.893	0.001	0.998
6.869	0.0005	0.999
11.18	0.00005	0.9999

$df = 6$

t	$P(T_6 \le -t)$	$P(-t \le T_6 \le t)$
1.440	0.1	0.8
1.943	0.05	0.9
2.447	0.025	0.95
3.143	0.01	0.98
3.707	0.005	0.99
5.208	0.001	0.998
5.959	0.0005	0.999
9.082	0.00005	0.9999

$df = 7$

t	$P(T_7 \le -t)$	$P(-t \le T_7 \le t)$
1.415	0.1	0.8
1.895	0.05	0.9
2.365	0.025	0.95
2.998	0.01	0.98
3.499	0.005	0.99
4.785	0.001	0.998
5.408	0.0005	0.999
7.885	0.00005	0.9999

$df = 8$

t	$P(T_8 \le -t)$	$P(-t \le T_8 \le t)$
1.397	0.1	0.8
1.860	0.05	0.9
2.306	0.025	0.95
2.896	0.01	0.98
3.355	0.005	0.99
4.501	0.001	0.998
5.041	0.0005	0.999
7.120	0.00005	0.9999

$df = 9$

t	$P(T_9 \le -t)$	$P(-t \le T_9 \le t)$
1.383	0.1	0.8
1.833	0.05	0.9
2.262	0.025	0.95
2.821	0.01	0.98
3.250	0.005	0.99
4.297	0.001	0.998
4.781	0.0005	0.999
6.594	0.00005	0.9999

$df = 10$

t	$P(T_{10} \le -t)$	$P(-t \le T_{10} \le t)$
1.415	0.1	0.8
1.895	0.05	0.9
2.365	0.025	0.95
2.998	0.01	0.98
3.499	0.005	0.99
4.785	0.001	0.998
5.408	0.0005	0.999
7.885	0.00005	0.9999

$df = 11$

t	$P(T_{11} \le -t)$	$P(-t \le T_{11} \le t)$
1.397	0.1	0.8
1.860	0.05	0.9
2.306	0.025	0.95
2.896	0.01	0.98
3.355	0.005	0.99
4.501	0.001	0.998
5.041	0.0005	0.999
7.120	0.00005	0.9999

$df = 12$

t	$P(T_{12} \le -t)$	$P(-t \le T_{12} \le t)$
1.383	0.1	0.8
1.833	0.05	0.9
2.262	0.025	0.95
2.821	0.01	0.98
3.250	0.005	0.99
4.297	0.001	0.998
4.781	0.0005	0.999
6.594	0.00005	0.9999

$df = 13$

t	$P(T_{13} \le -t)$	$P(-t \le T_{13} \le t)$
1.350	0.1	0.8
1.771	0.05	0.9
2.160	0.025	0.95
2.650	0.01	0.98
3.012	0.005	0.99
3.852	0.001	0.998
4.221	0.0005	0.999
5.513	0.00005	0.9999

$df = 14$

t	$P(T_{14} \le -t)$	$P(-t \le T_{14} \le t)$
1.345	0.1	0.8
1.761	0.05	0.9
2.145	0.025	0.95
2.624	0.01	0.98
2.977	0.005	0.99
3.787	0.001	0.998
4.140	0.0005	0.999
5.363	0.00005	0.9999

$df = 15$

t	$P(T_{15} \le -t)$	$P(-t \le T_{15} \le t)$
1.341	0.1	0.8
1.753	0.05	0.9
2.131	0.025	0.95
2.602	0.01	0.98
2.947	0.005	0.99
3.733	0.001	0.998
4.073	0.0005	0.999
5.239	0.00005	0.9999

$df = 16$

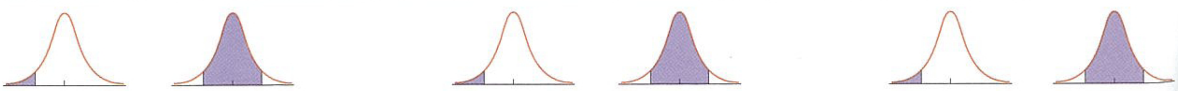
t	$P(T_{16} \le -t)$	$P(-t \le T_{16} \le t)$
1.337	0.1	0.8
1.746	0.05	0.9
2.120	0.025	0.95
2.583	0.01	0.98
2.921	0.005	0.99
3.686	0.001	0.998
4.015	0.0005	0.999
5.134	0.00005	0.9999

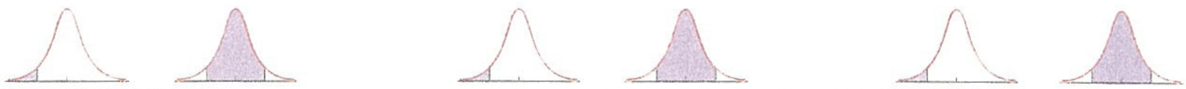
$df = 17$

t	$P(T_{17} \le -t)$	$P(-t \le T_{17} \le t)$
1.333	0.1	0.8
1.740	0.05	0.9
2.110	0.025	0.95
2.567	0.01	0.98
2.898	0.005	0.99
3.646	0.001	0.998
3.965	0.0005	0.999
5.044	0.00005	0.9999

$df = 18$

t	$P(T_{18} \le -t)$	$P(-t \le T_{18} \le t)$
1.33	0.1	0.8
1.734	0.05	0.9
2.101	0.025	0.95
2.552	0.01	0.98
2.878	0.005	0.99
3.610	0.001	0.998
3.922	0.0005	0.999
4.966	0.00005	0.9999





$df = 19$

t	$P(T_{19} \leq -t)$	$P(-t \leq T_{19} \leq t)$
1.328	0.1	0.8
1.729	0.05	0.9
2.093	0.025	0.95
2.539	0.01	0.98
2.861	0.005	0.99
3.579	0.001	0.998
3.883	0.0005	0.999
4.897	0.00005	0.9999

$df = 20$

t	$P(T_{20} \leq -t)$	$P(-t \leq T_{20} \leq t)$
1.325	0.1	0.8
1.725	0.05	0.9
2.086	0.025	0.95
2.528	0.01	0.98
2.845	0.005	0.99
3.552	0.001	0.998
3.850	0.0005	0.999
4.837	0.00005	0.9999

$df = 22$

t	$P(T_{22} \leq -t)$	$P(-t \leq T_{22} \leq t)$
1.321	0.1	0.8
1.717	0.05	0.9
2.074	0.025	0.95
2.508	0.01	0.98
2.819	0.005	0.99
3.505	0.001	0.998
3.792	0.0005	0.999
4.736	0.00005	0.9999

$df = 24$

t	$P(T_{24} \leq -t)$	$P(-t \leq T_{24} \leq t)$
1.318	0.1	0.8
1.711	0.05	0.9
2.064	0.025	0.95
2.492	0.01	0.98
2.797	0.005	0.99
3.467	0.001	0.998
3.745	0.0005	0.999
4.654	0.00005	0.9999

$df = 26$

t	$P(T_{26} \leq -t)$	$P(-t \leq T_{26} \leq t)$
1.315	0.1	0.8
1.706	0.05	0.9
2.056	0.025	0.95
2.479	0.01	0.98
2.779	0.005	0.99
3.435	0.001	0.998
3.707	0.0005	0.999
4.587	0.00005	0.9999

$df = 28$

t	$P(T_{28} \leq -t)$	$P(-t \leq T_{28} \leq t)$
1.313	0.1	0.8
1.701	0.05	0.9
2.048	0.025	0.95
2.467	0.01	0.98
2.763	0.005	0.99
3.408	0.001	0.998
3.674	0.0005	0.999
4.530	0.00005	0.9999

$df = 30$

t	$P(T_{30} \leq -t)$	$P(-t \leq T_{30} \leq t)$
1.31	0.1	0.8
1.697	0.05	0.9
2.042	0.025	0.95
2.457	0.01	0.98
2.75	0.005	0.99
3.385	0.001	0.998
3.646	0.0005	0.999
4.482	0.00005	0.9999

$df = 32$

t	$P(T_{32} \leq -t)$	$P(-t \leq T_{32} \leq t)$
1.309	0.1	0.8
1.694	0.05	0.9
2.037	0.025	0.95
2.449	0.01	0.98
2.738	0.005	0.99
3.365	0.001	0.998
3.622	0.0005	0.999
4.441	0.00005	0.9999

$df = 34$

t	$P(T_{34} \leq -t)$	$P(-t \leq T_{34} \leq t)$
1.307	0.1	0.8
1.691	0.05	0.9
2.032	0.025	0.95
2.441	0.01	0.98
2.728	0.005	0.99
3.348	0.001	0.998
3.601	0.0005	0.999
4.405	0.00005	0.9999

$df = 36$

t	$P(T_{36} \leq -t)$	$P(-t \leq T_{36} \leq t)$
1.306	0.1	0.8
1.688	0.05	0.9
2.028	0.025	0.95
2.434	0.01	0.98
2.719	0.005	0.99
3.333	0.001	0.998
3.582	0.0005	0.999
4.374	0.00005	0.9999

$df = 40$

t	$P(T_{40} \leq -t)$	$P(-t \leq T_{40} \leq t)$
1.303	0.1	0.8
1.684	0.05	0.9
2.021	0.025	0.95
2.423	0.01	0.98
2.704	0.005	0.99
3.307	0.001	0.998
3.551	0.0005	0.999
4.321	0.00005	0.9999

$df = 50$

t	$P(T_{50} \leq -t)$	$P(-t \leq T_{50} \leq t)$
1.299	0.1	0.8
1.676	0.05	0.9
2.009	0.025	0.95
2.403	0.01	0.98
2.678	0.005	0.99
3.261	0.001	0.998
3.496	0.0005	0.999
4.228	0.00005	0.9999

$df = 60$

t	$P(T_{60} \leq -t)$	$P(-t \leq T_{60} \leq t)$
1.296	0.1	0.8
1.671	0.05	0.9
2.000	0.025	0.95
2.390	0.01	0.98
2.660	0.005	0.99
3.232	0.001	0.998
3.460	0.0005	0.999
4.169	0.00005	0.9999

$df = 75$

t	$P(T_{75} \leq -t)$	$P(-t \leq T_{75} \leq t)$
1.293	0.1	0.8
1.665	0.05	0.9
1.992	0.025	0.95
2.377	0.01	0.98
2.643	0.005	0.99
3.202	0.001	0.998
3.425	0.0005	0.999
4.110	0.00005	0.9999

$df = 100$

t	$P(T_{100} \leq -t)$	$P(-t \leq T_{100} \leq t)$
1.290	0.1	0.8
1.660	0.05	0.9
1.984	0.025	0.95
2.364	0.01	0.98
2.626	0.005	0.99
3.174	0.001	0.998
3.390	0.0005	0.999
4.053	0.00005	0.9999

$df = 125$

t	$P(T_{125} \leq -t)$	$P(-t \leq T_{125} \leq t)$
1.288	0.1	0.8
1.657	0.05	0.9
1.979	0.025	0.95
2.357	0.01	0.98
2.616	0.005	0.99
3.157	0.001	0.998
3.370	0.0005	0.999
4.020	0.00005	0.9999

$df = 150$

t	$P(T_{150} \leq -t)$	$P(-t \leq T_{150} \leq t)$
1.287	0.1	0.8
1.655	0.05	0.9
1.976	0.025	0.95
2.351	0.01	0.98
2.609	0.005	0.99
3.145	0.001	0.998
3.357	0.0005	0.999
3.998	0.00005	0.9999

$df = \infty$

t	$P(Z \leq -t)$	$P(-t \leq Z \leq t)$
1.282	0.1	0.8
1.645	0.05	0.9
1.960	0.025	0.95
2.326	0.01	0.98
2.576	0.005	0.99
3.090	0.001	0.998
3.291	0.0005	0.999
3.891	0.00005	0.9999